HYPOTHESES ON THE SEASONAL DISTRIBUTION OF THUNDERSTORM RAINFALL IN SOUTHEASTERN ARIZONA

H153

By

Leonard J. Lane
and
Herbert B. Osborn
Hydrologist and Hydraulic Engineer, respectively
Southwest Watershed Research Center
Agricultural Research Service
U.S. Department of Agriculture
442 E. 7th Street
Tucson, Arizona 85705

Abstract

In much of southern Arizona, thunderstorm rainfall is the most important source of water. Thus, estimates of thunderstorm rainfall, or the lack of it, are important in decisions concerning water supplies and water resource system design. Observations have confirmed recent analyses that the extreme flood events on small (100 square miles and less) semiarid watersheds in southern Arizona are due to air-mass thunderstorm rainfall. Although there is no accurate way to forecast seasonal rainfall, knowledge of thunderstorm rainfall frequencies is essential for evaluating water resource systems. If data are available to aid in decisions, they usually are point measurements from a single station. In this paper, analysis indicates that a single gage is not sufficient to measure thunderstorm rainfall on the 58-square-mile Walnut Gulch watershed.

The occurrence of significant storms during the summer rainy season is modeled as a Bernoulli random variable with parameter p_i varying within the season. Of interest is the number of rainy days during the season and during shorter periods of time within the season.

Equations are presented for the mean and variance of the number of rainy days, N_r, for two hypotheses concerning the variability of the Bernoulli parameters, p_i, during the season. The first hypothesis is that the p_i are constant throughout the variable season. In this case, the distribution of N_r is compound, and the distributions of the number of events in fixed subintervals is binomial. Under the second hypothesis, there is strong seasonality with considerable variation of the p_i. Equations, in terms of means and variances under the two hypotheses, are provided to make preliminary tests of those hypotheses. Short records proved inadequate to decide in favor of either hypothesis.

Proc. Second Int. Symp. in Hydrology, (Sept. 1972) Ft. Collins, Water Res. Publication Ft. Collins, 1973: 83-94.

A simulated distribution is compared to the normal distribution to show what characteristics of the p_i are useful in choosing the best probability distribution to approximate N_r under the second hypothesis. The normal distribution is shown to be an excellent approximation to the distribution of the number of events per season.

Introduction

Since semiarid regions represent natural systems that tend to be fragile and are undergoing change [1], information is needed as a basis for a range in choice in future hydrologic development of these lands. Such information is needed to enhance utilization of the natural resources of the arid and semiarid lands as these lands are submitted to greater stress from an increasing population. An integral part of this needed information is knowledge of thunderstorm rainfall characteristics on semiarid watersheds.

In analyses of convective storms in the Southwest, Osborn and Reynolds [2] found that thunderstorm rainfall is a most important source of moisture in the Southwest. They characterized convective rainfall as intense, short-duration rainfall of limited areal extent. In their words, "Such storms cause most of the floodwater damage, surface erosion, arroyo formation, and sediment deposition. They also contribute occasionally to flows into reservoir storage for downstream use. Because of the importance of such ephemeral stream flows in the intermountain and high plains areas of the Southwest, knowledge of frequency of storms that produce them is important:" (p. 81).

Distribution and significance of thunderstorm rainfall vary regionally and topographically throughout the Southwest. In this paper, once the seasonal distribution of convective storms in Arizona is established [3], the occurrence of significant storms during the summer rainy season on the Walnut Gulch Experimental Watershed in southeastern Arizona is modeled as a Bernoulli random variable with two assumptions about the parameter. Of interest are the number of occurrences of rainy days during the season and during shorter periods within the season.

Characteristics of Thunderstorm Rainfall

Annual precipitation in southeastern Arizona ranges from less than 10 inches in lower valleys to over 25 inches in some mountainous areas. The 68-year mean annual rainfall at Tombstone, Arizona (near the center of the Walnut Gulch watershed) is 14.0 inches [4]. There is a pronounced seasonal distribution of rainfall in southeastern Arizona, with precipitation in the winter characterized by low intensity frontal rainfall and in the summer by air-mass thunderstorms [3,5].

Monthly distribution of precipitation for Tombstone, Arizona is shown in Fig. 1. The average number of days with 0.10 inch or more of precipitation is shown in Fig. 1a, while the corresponding monthly percentages of the annual total are shown in Fig. 1b. Roughly half of the days with 0.10 inch or more are in July, August, and September, and over 60% of the annual precipitation occurs in these three months.

Thunderstorm rainfall exhibits a pronounced diurnal effect[6] with maximum activity in the afternoon or early evening. Observations of thunderstorms in the Tombstone area indicate a mean starting time of about 5 p.m., with about 2/3 of the storms starting within a 6- to 7-hour period centered about 5 p.m.

The Walnut Gulch Experimental Watershed, as described by Renard [7], is a 58-square-mile semiarid rangeland area surrounding Tombstone, Arizona. There are 98 recording rain gages with various record lengths on or immediately adjacent to the experimental watershed. Average annual precipitation (Thiessen weighted) over the watershed and the minimum and maximum point rainfall for each of the years 1954-1969 (Fig. 2) illustrates the variability of point and average precipitation. Additional references to the variability of thunderstorm rainfall in individual storms for Walnut Gulch are Renard [8] and Osborn, Lane, and Hundley [9] and for a smaller experimental rangeland area near Tucson, Arizona, are Woolhiser and Schwalen [10] and Fogel and Duckstein [11].

Seasonal Distribution of Thunderstorm Rainfall

Two major problems in modeling thunderstorm rainfall are modeling the occurrence of storms on a day within the rainy season and modeling the dynamic thunderstorm processes. The first problem is by far simpler.

Of primary importance in defining a summer rainy season are the physical relationships governing commencement of the rainy season. Thunderstorm development is affected by a host of meteorological variables; for example, amount and distribution of moist air aloft, temperature distributions, winds, etc. Sellers [3] describes the typical weather patterns associated with thunderstorm development in Arizona. A singularly important factor is the establishment of sea level and 500 mb pressure patterns enabling the moist air flow from the Gulf of Mexico. Typically, the "correct" pressure patterns are established in early summer, producing a predominance of thunderstorm rainfall in July and August (Fig. 1).

Gabriel and Neumann [12] state that the Markov chain model for daily rainfall had been suggested as early as 1953 by Brooks and Carruthers [13]. Such a model was fitted using 27 years of data for Tel Aviv by Gabriel and Neumann, who found

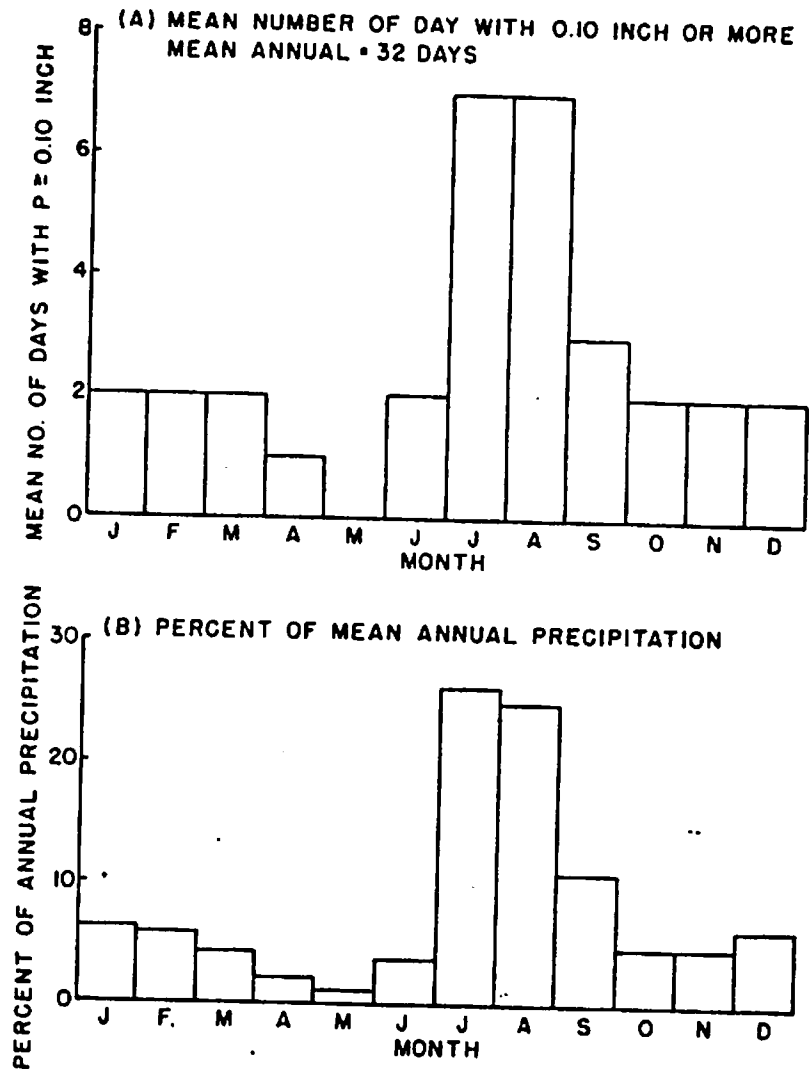

Fig. 1. Monthly distribution of precipitation for Tombstone, Arizona (from Green and Sellers, Arizona Climate, University of Arizona Press, 1964).

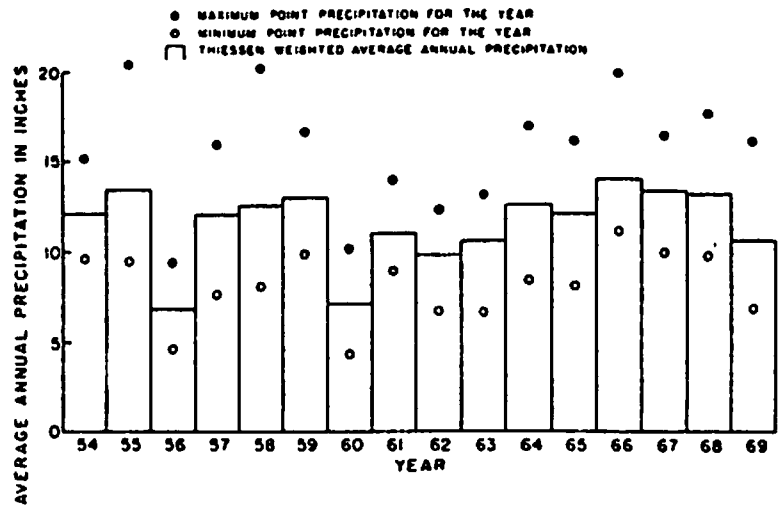

Fig. 2a. Walnut Gulch average annual precipitation over the watershed 1954-1969 mean = 11.5 inches.

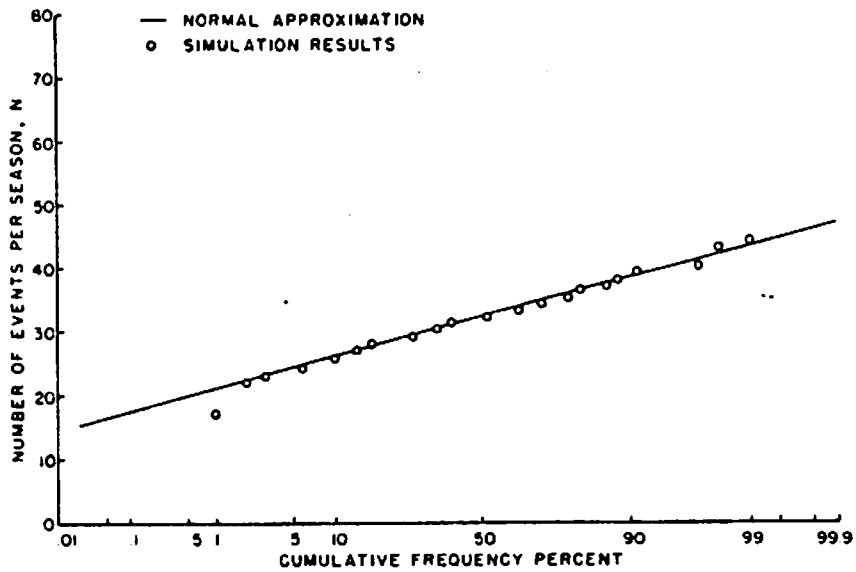

Fig. 2b. Number of events per season in 100 years synthetic data for the Walnut Gulch watershed.

a seasonal trend in the probabilities. Analysis of these data led the authors to conclude that Tel Aviv rainfall events were independent of rainfall more than 5 days previous to the given day. Analysis by Caskey [14] supported the choice of a Markov chain model for precipitation at Denver, Colorado. In an analysis of rainfall data throughout the United States, Hershfield [15] computed relationships between conditional probabilities and unconditional probabilities for various thresholds on wet and dry sequences. A strong relationship was found between the unconditional probabilities and conditional probabilities for dry sequences. This correspondence increased with the threshold values, suggesting that for threshold values of sufficient magnitude the probabilities are equal. For the wet sequences, the relationship between the probabilities seemed to decrease with increasing threshold. The analysis of rainfall data presented herein assumes the Bernoulli distribution for daily rainfall derived from an assumption other than the Markov chain model. Gabriel and Neumann [12] recognized that the derivation of the geometric distributions for wet and dry spells could arise from other than the Markov chain model. This paper presents two alternate hypotheses suggesting such a derivation.

A previous attempt at modeling the probability of a thunderstorm occurring during the summer rainy season involved assuming a probability distribution for the start of the rainy season. When the season started, the occurrence of an event was modeled as a Bernoulli random variable with constant parameter throughout the season. Alternate hypotheses are that the Bernoulli parameter is constant but the length of season varies, and that the Bernoulli parameter varies within the season.

In the present analyses, persistence is assumed to exist, but is difficult to measure and model. The following statement was included in U.S. Weather Bureau Technical Paper 38 [16]: "While there is a tendency for persistence in weather, major storms are rare events that, in a sense, may be regarded as exceptions to persistence. Therefore, while it is physically possible for outstanding storms to repeat at intervals of something like 4 to 5 days, there is relatively little probability of the occurrence of such a series" (p. 52). Therefore, in these analyses, the existence of persistence is recognized but not modeled directly. The first hypothesis is that the occurrence rate within the season is constant once the season has begun. The length of the season is variable, producing a preponderance of events near the middle of the season when several years of data are analyzed. Assertions under this hypothesis are: 1) that the occurrence of storms can be modeled as a Bernoulli variable with constant parameter p throughout the season; 2) that the number of rainy days per season for fixed season length is a binomial random variable; 3) that the number of rainy days for a variable length of season is compound; and 4) with these distributional assumptions

that the variance of the number of events per season must exceed the mean number of events. The second hypothesis is that the occurrence rate within the season is variable. Assertions under this hypothesis are that the occurrence of storms can be modeled as a Bernoulli random variable with varying parameter p_k. From this assumption, the variance of the number of events must be less than the mean number of events [17].

The occurrence of a significant storm (depth equal to or exceeding 0.25 inch) is modeled as a Bernoulli random variable. While the threshold or critical value of 0.25 inch is arbitrary, there is empirical evidence to support it [18]. Different values for the threshold constant would affect the statistics, but this would not be expected to change the hypothesis of a Bernoulli random variable.

The Bernoulli random variable, X_i, is 1 if there is a significant rain on day i, and 0 otherwise:

$$X_i = \begin{cases} 1 & \text{if significant rain on day } i \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

such that the probability of significant rain on day i is:

$$P(X_i = 1) = p_i \text{ , and} \tag{2}$$

$$P(X_i = 0) = q_i = 1 - p_i . \tag{3}$$

If the random variable N_r is defined as the number of rainy days during the season, then:

$$N_r = \sum_{i=1}^{n} X_i \quad , \tag{4}$$

where the X_i are Bernoulli random variables as defined above, and n is the number of days in the season. The expected value of N_r is then the sum of the expectations for the X_i.

$$E(N_r) = \sum_{i=1}^{n} p_i \quad , \tag{5}$$

where E denotes expectation. Furthermore, if we assume no persistence, then the X_i are independent, and

$$Var(N_r) = \sum_{i=1}^{n} p_i q_i = \sum_{i=1}^{n} p_i - \sum_{i=1}^{n} p_i^2 \quad . \tag{6}$$

where Var denotes variance.

If all the p_i are equal to p, then the number of rainy days (for fixed n) has a binomial distribution with parameters n and p [17]. In this case, we have:

$$E(N_r) = \sum_{i=1}^{n} p = np \ , \tag{7}$$

and

$$Var(N_r) = \sum_{i=1}^{n} pq = npq \ . \tag{8}$$

Since $p + q = 1$, we must have

$$Var(N_r) < E(N_r) \ , \tag{9}$$

because $Var(N_r) = q \cdot E(N_r)$. (10)

The first hypothesis is that the length of season, N_s , is a random variable but that the parameters p_i are all equal to p . Under this assumption, for the number of rainy days:

$$E(N_r) = E(N_s)\ E(X) \ , \tag{11}$$

and

$$Var(N_r) = E(N_s)\ Var(X) + Var(N_s)\ [E(X)]^2 \ . \tag{12}$$

A preliminary test of this hypothesis can be made by considering the ratio of the variance to the mean:

$$Var(N_r)/E(N_r) = (1-p) + p[Var(N_s)/E(N_s)] \ , \tag{13}$$

where the variables are as explained above.

If we let $n \to \infty$ and all $p_i \to 0$ in such a way that $np \to u$, then Eq. 4 specifies a Poisson distribution with

$$E(N_r) = u \ , \tag{14}$$

and

$$Var(N_r) = u \ . \tag{15}$$

Feller [17] proves this by showing that the limiting form of the generating function for N_r specifies a Poisson distributic

The last case we consider is for finite n and unequal p_i . In this case Eqs. 5 and 6 apply. As stated earlier, the variance as calculated by Eq. 6 is maximum when all the p_i are equal. Thus, we know that for the p_i not all equal:

$$Var(N_r) < npq < E(N_r) \ , \tag{16}$$

where p is the "average" of the p_i ; that is,

$$p = (\sum_{i=1}^{n} p_i)/n \ . \tag{17}$$

If the p_i are all equal, then they can be thought of as being uniformly distributed within the season. The first

hypothesis is that the occurrence rate within the season is constant once the season has begun, but that the length of the season is variable, producing a preponderance of events near the middle of the season in the aggregate curve. The second hypothesis is that the occurrence rate within the individual seasons is variable and the beginning of the seasons may be variable. If we can neglect the variability in the beginning of the seasons, then Eqs. 11, 12, and 13 apply under the first hypothesis, and under the second hypothesis Eqs. 5, 6, and 16 apply. Moreover, Eqs. 13 and 16 provide a preliminary test of the two hypotheses.

Results and Data

On the Walnut Gulch watershed, 47 recording rain gages have continuous records from 1960 to 1970. The number of significant storms (0.25 inch or greater) per summer rainy season was tabulated for several rain gage networks. The first network had all 47 gages, the next had 8 gages evenly spaced over the 57-square-mile watershed, and the last "network" consisted of a single gage chosen at random. Data from these networks are shown in Table 1. Notice that the means and variances for the areal measurements should be larger than for point measurements since they sample in time and in space. That is, the events in space are not completely dependent, so that their variances are constructed in an additive manner. The fact that the mean number of rainy days is much smaller for the single gage network than for the 47-gage network supports the observations that a single or "key" gage is not adequate for thunderstorm rainfall measurement on Walnut Gulch [19].

Table 1. Statistics on the number of rainy days (0.25 inch or greater) per season, Walnut Gulch, 1960-1970.

Network	Number of Gages	Number of rainy days per season Mean	Variance
1	47*	35.0	47.6
2	8*	27.2	29.8
3	1**	10.5	7.47

*Nearly uniform over the watershed area
**Chosen at random, RG-27

As a preliminary test of the alternate hypotheses on the variability of the Bernoulli parameters, ratios of the variance to the mean number of rainy days were computed as shown in Table 2. Again, hypothesis 1 assumes a variable season length with constant p. These ratios are calculated using Eq. 13. For the second hypothesis, an upper limit to the ratios is calculated using Eq. 10. The last column in Table 2 shows the

ratios for the observed data. Except for the 1-gage network, these ratios do not support one hypothesis over the other.

Table 2. Ratios of the variance to the mean for the number of rainy days per season.

Network	Number of gages	Ratio of Variance to Mean: Hypothesis-1 constant p	Hypothesis-2 variable p*	Observed data
1	47	2.26	0.77	1.36
2	8	2.98	0.82	1.10
3	1	1.27	0.93	0.71

*Upper limit for this ratio, actual values would be less than shown.

An empirically derived curve starting with $p_1 = 0.0$ on May 15, a maximum of $p_{72} = 0.51$ on July 25, and $p_{154} = 0.0$ on October 15 was derived from 10 years of data on Walnut Gulch. This curve, which specifies the variable Bernoulli parameters throughout the season, was used to generate synthetic rainfall data on Walnut Gulch. The expected number of events from this curve of probabilities is 32.4 (using Eq. 5) with a variance of 19.9 calculated using Eq. 6. Using the curve of probabilities, 100 years of synthetic data were produced for Walnut Gulch. Although the mean value differs from that shown for the 47-gage network in Table 1, these synthetic data were used to judge the normal approximation to the number of events per season. This approximation to the synthetic data is shown in Fig. 2b. As indicated by the central limit theorem, if the variable parameter Bernoulli model is acceptable, the normal distribution is an excellent approximation to the number of events per season.

Summary and Conclusions

Alternate hypotheses are proposed for modeling the occurrence of thunderstorm rainfall during a summer rainy season in southeastern Arizona. Physical considerations seem to favor the variable parameter Bernoulli variable, but means and variances calculated from short records do not favor one hypothesis over the other.

Mean values for the number of rainy days per season for 11 years of record on the 58-square-mile Walnut Gulch watershed increase with the number of gages in the network. This evidence supports previous conclusions that the "key-gage" theory is incorrect for air-mass thunderstorm rainfall in southeastern Arizona.

If the variable parameter Bernoulli model is acceptable for air-mass thunderstorm rainfall in southeastern Arizona, then the normal distribution can be used to approximate the distribution of the number of events per season.

Further studies are underway to define better the point-area relationship for thunderstorm rainfall on Walnut Gulch and to compare these results to other point and watershed rainfall data in the Southwest.

References

[1] Hastings, J. R. and R. M. Turner, The Changing Mile. The University of Arizona Press, Tucson, Arizona, (1965).

[2] Osborn, H. B. and W. N. Reynolds, Convective Storm Patterns in the Southwestern United States. Bull. IASH 8(3): 71-83, (1963).

[3] Green, C. R. and W. D. Sellers, Eds. Arizona Climate, The University of Arizona Press, Tucson, Arizona, (1964).

[4] Hydrologic Data for Experimental Agricultural Watersheds in the United States, 1964. Compiled by J. B. Burford, USDA Misc. Pub. No. 1194, (June, 1971).

[5] Osborn, H. B., L. J. Lane, and R. S. Kagan, Stochastic Models of Spatial and Temporal Distribution of Thunderstorm Rainfall. Proceedings, USDA-IASPS Symposium on Statistical Hydrology, Tucson, Arizona, (1971).

[6] Petterssen, S., Weather Analysis and Forecasting, McGraw-Hill Book Co., Inc., New York, Vol. 2, (1956).

[7] Renard, K. G., The Hydrology of Semiarid Rangeland Watersheds. USDA, Pub. ARS 41-162, 26 pp. (1970).

[8] Renard, K. G., ibid, (1970).

[9] Osborn, H. B., L. J. Lane, and J. F. Hundley, Optimum Gaging of Thunderstorm Rainfall in Southeastern Arizona. Water Resources Research, AGU 8(1): 259-265, (1972).

[10] Woolhiser, D. A. and H. A. Schwalen, Area-Depth Frequency Relations for Thunderstorm Rainfall in Southern Arizona. University of Arizona, Agr. Exp. Station Tech. Paper 527, (1959).

[11] Fogel, M. M. and L. Duckstein, Point Rainfall Frequencies in Convective Storms. Water Resources Res. 5(6):1229-1237, (1969).

[12] Gabriel, K. R., and J. Neumann, A Markov Chain Model for Daily Rainfall Occurrence at Tel Aviv. Quarterly Journal of the Royal Meteorological Society, 88(375) pp. 90-95, (January, 1962).

[13] Brooks, C. E. P., and N. C. Carruthers, Handbook of Statistical Methods in Meteorology, London, H.M.S.O. (1953).

[14] Caskey, J. E., A Markov Chain Model for the Probability of Precipitation Occurrence in Intervals of Various Length, Monthly Weather Review, pp. 298-301, (June 1963).

[15] Hershfield, D. M., A Comparison of Conditional and Unconditional Probabilities for Wet- and Dry-Day Sequences, J. of Applied Meteorology, 9(5), pp. 825-827, (October 1970).

[16] Technical Paper No. 38, USWB, U.S. Dept. of Commerce, Washington, D. C., (1960).

[17] Feller, W. An Introduction to Probability Theory and Its Applications, Vol. I, John Wiley & Sons, Inc., New York, (1950).

[18] Osborn, H. B., and L. Lane, Precipitation-Runoff Relationships for Very Small Semiarid Rangeland Watersheds, Water Resources Res., AGU 5 (2):419-425, (1969).

[19] Hershfield, D. M., Rainfall Input for Hydrologic Models, Estract of Geochemistry, Precipitation, Evaporation, Soil-Moisture, Hydrometry, General Assembly of Bern, (September-October, 1967).